Brownian Confidence Bands on Monte Carlo Output

W.S. Kendall, Jean-Michel Marin, C.P. Robert

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Brownian Confidence Bands on Monte Carlo Output

W.S. Kendall — J.M. Marin — C.P. Robert

N° 5436

Décembre 2004

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--5436--FR+ENG

Brownian Confidence Bands on Monte Carlo Output

W.S. Kendall* , J.M. Marin† , C.P. Robert‡

Thème COG — Systèmes cognitifs
Projets Select

Rapport de recherche n° 5436 — Décembre 2004 — 16 pages

Abstract: When considering a Monte Carlo estimation procedure, the path produced by successive partial estimates is often used as a guide for informal convergence diagnostics. However the confidence region associated with that path cannot be derived simplistically from the confidence interval for the estimate itself. An asymptotically correct approach can be based on the Brownian motion approximation of the path, but no exact formula for the corresponding area-minimizing confidence region is yet known. We construct proxy regions based on local time arguments and consider numerical approximations. These are then available for a more incisive assessment of the Monte Carlo procedure and thence of the estimate itself.

Key-words: Boundary crossing probability, Brownian motion, central limit theorem, local time, Monte Carlo path, prediction region, simultaneous confidence region

* Department of Statistics, University of Warwick, Coventry CV4 7AL, UK
† Project SELECT, INRIA FUTURS, University Paris-Sud, Orsay and CEREMADE, University Paris Dauphine, France
‡ CEREMADE, University Paris Dauphine and CREST, INSEE, Paris, France

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,
4, rue Jacques Monod, 91893 ORSAY Cedex (France)
Téléphone : +33 1 72 92 59 00 —Télécopie : +33 1 72 92 59 ??

Bandes de confiance browniennes pour trajectoires de Monte-Carlo

Résumé : Lorsque l'on approxime une intégrale par la méthode de Monte-Carlo, il est d'usage de représenter la trajectoire des estimations partielles, dite trajectoire de Monte-Carlo. Cela permet de visualiser la qualité de l'approximation obtenue. Contrairement à une pratique très répandue, il ne nous semble pas opportun de représenter également la région définie par les intervalles de confiance associés aux estimations partielles. En effet, la région de confiance ainsi obtenue n'est pas simultanée et ne porte pas directement sur la trajectoire de Monte-Carlo. Nous proposons de tracer à la place une région de prédiction simultanée, i.e. une surface contenant toute nouvelle trajectoire avec une probabilité fixée. Pour ce faire, nous utilisons une approche asymptotique et la convergence en loi des trajectoires vers un mouvement brownien. Enfin, nous traitons du problème lié à la détermination de la région de prédiction optimale, i.e. de surface minimale.

Mots-clés : Trajectoires de Monte-Carlo, région de prédiction, mouvement brownien, théorème central limite, temps local

1 Introduction

The Monte Carlo method (see, for example, Robert and Casella, 2004) is validated by both the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT):

(a) the (almost sure) convergence of an average

$$\widehat{\mathfrak{I}}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \quad \longrightarrow \quad \mathfrak{I} = \int h(x)\pi(\mathrm{d}x) \tag{1}$$

is guaranteed by the LLN when the X_i's are drawn independently from the probability measure π and h is integrable against π;

(b) the variation of $\widehat{\mathfrak{I}}_n - \mathfrak{I}$ is asymptotically Gaussian when h^2 is integrable against π.

There is however a difficulty in using the CLT to assess the convergence of the sequence $(\widehat{\mathfrak{I}}_n)_{1\le n\le N}$: while the CLT tells us that the limiting value $\mathfrak{I}$ lies in the interval $[\widehat{\mathfrak{I}}_N - 2\sigma/\sqrt{N}, \widehat{\mathfrak{I}}_N + 2\sigma/\sqrt{N}]$, with probability close to 0.95 for large enough N, it does not provide information on the probabilistic behaviour of the *sequence* $(\widehat{\mathfrak{I}}_n)_{1\le n\le N}$. In other words, the CLT is not helpful in assessing where the whole *path* $(\widehat{\mathfrak{I}}_n)_{1\le n\le N}$ might lie, were we to start a new simulation experiment. Nevertheless, the shape of $(\widehat{\mathfrak{I}}_n)_{1\le n\le N}$ is customarily used to decide about the stability of a Monte Carlo estimate and the relevance of the corresponding Monte Carlo experiment. In a fairly informal way, we often decide to stop a Monte Carlo experiment, (which is to say, we choose the final value of N), and to adopt the final value of $\widehat{\mathfrak{I}}_N$ as an acceptable approximation to $\mathfrak{I}$, based on the appearance of $(\widehat{\mathfrak{I}}_n)_{1\le n\le N}$, even though another run of the Monte Carlo experiment, based on a different pseudo-random seed, might have produced a very different appearance and thus might have led to a different conclusion (to wit, to continue the Monte Carlo experiment and use more simulations). It thus seems necessary to provide a confidence assessment on the entire trajectory of the Monte Carlo sequence so as to help to quantify this kind of decision.

Therefore we need to approximate the trajectory of $(\widehat{\mathfrak{I}}_n)_{1\le n\le N}$ in order to see how to assess its variability as a sequence without running an extensive Monte Carlo experiment that would require many replications of $(\widehat{\mathfrak{I}}_n)_{1\le n\le N}$. Such an approximation is found in Donsker's theorem, and relates the sequence to a Brownian motion $\{W(t)\}_{0\le t\le 1}$. We are therefore looking for a "smallest" region $\mathscr{R} \subset [0,1]\times\mathbb{R}$ that contains $\{W(t)\}_{0\le t\le 1}$ with a high enough probability. The purpose of this paper is to derive such a region $\mathscr{R}$, based on a minimal area optimality criterion, and to study its use in Monte Carlo experiments. In Section 2, we restate the problem in more formal terms; in Sections 3 and 4, we derive quasi-optimal confidence and prediction regions, while Section 5 illustrates the confidence procedures in a specific Monte Carlo setting and Section 6 concludes the paper.

2 Simultaneous confidence and prediction regions

To formalise the setup further, consider π a probability distribution and the quantity of interest,

$$\pi(h) = \int h(x)\pi(\mathrm{d}x) = \mathbb{E}_\pi[h(X)]\,,$$

approximated by the standard Monte Carlo estimator $\widehat{\mathfrak{I}}_n$ given by (1). The sequence $(\widehat{\mathfrak{I}}_n)_{1\le n\le N}$ is converted to the random step function

$$\{\delta_N(t)\}_{t\in[0,1]} = \left(\frac{1}{\lfloor Nt\rfloor}\sum_{i=1}^{\lfloor Nt\rfloor} h(X_i)\right)_{t\in[0,1]}$$

where $\lfloor x\rfloor$ denotes the integer part of x (the greatest integer not exceeding x). Note: $\delta_N(1) = \widehat{\mathfrak{I}}_N$.

2.1 Confidence regions

To fix ideas, we first reformulate the CLT in this setting: for a fixed $t\in[0,1]$, if $\mathbb{E}_\pi[|h(X)|^2] < \infty$ then the random variable

$$\sqrt{\lfloor Nt\rfloor}\left(\frac{\delta_N(t) - \pi(h)}{\mathbb{V}_\pi[h(X)]^{1/2}}\right)$$

converges in distribution to a standard $\mathcal{N}(0,1)$ random variable, where

$$\mathbb{V}_\pi[h(X)] = \int h^2(x)\pi(\mathrm{d}x) - \left(\int h(x)\pi(\mathrm{d}x)\right)^2 .$$

In a standard Monte Carlo setting, it is customary to replace $\mathbb{V}_\pi[h(X)]$ with an estimate. Consequently we introduce the random function

$$\{\gamma_N(t)\}_{t\in[0,1]} = \left(\frac{1}{\lfloor Nt \rfloor} \sum_{i=1}^{\lfloor Nt \rfloor} (h(X_i) - \delta_N(t))^2 \right)_{t\in[0,1]} ;$$

if $\mathbb{E}_\pi[|h(X)|^2] < \infty$ and $t \in [0,1]$ is fixed then $\gamma_N(t)$ almost surely converges to $\mathbb{V}_\pi[h(X)]$ and the random variable

$$\sqrt{\lfloor Nt \rfloor} \left(\frac{\delta_N(t) - \pi(h)}{\gamma_N(t)^{1/2}} \right)$$

also converges in distribution to a standard normal random variable by virtue of Slutsky's theorem (Billingsley, 1995).

Therefore if we fix both $t \in [0,1]$ and $\alpha \in]0, 0.5[$ then the (random) confidence interval

$$\left[\delta_N(t) - \frac{\sqrt{\gamma_N(t)}}{\sqrt{\lfloor Nt \rfloor}} \Phi^{-1}(1-\alpha/2) \, , \; \delta_N(t) + \frac{\sqrt{\gamma_N(t)}}{\sqrt{\lfloor Nt \rfloor}} \Phi^{-1}(1-\alpha/2) \right] \tag{2}$$

contains $\pi(h)$ with a probability converging to $1-\alpha$ when N goes to infinity. This confidence interval is thus justified from probability theory and it is appropriately used in simulation experiments as an indicator of the precision of the Monte Carlo estimator.

However, if we use the interval (2) for each value of $t \in [0,1]$ then the confidence band thus built around $\{\delta_N(t)\}_{t\in[0,1]}$,

$$\left\{ (t,y) \in [0,1] \times \mathbb{R} \;\middle|\; y \in \left[\delta_N(t) - \frac{\sqrt{\gamma_N(t)}}{\sqrt{\lfloor Nt \rfloor}} \Phi^{-1}(1-\alpha/2), \delta_N(t) + \frac{\sqrt{\gamma_N(t)}}{\sqrt{\lfloor Nt \rfloor}} \Phi^{-1}(1-\alpha/2) \right] \right\} , \tag{3}$$

is not valid. Indeed, this is *not* a simultaneous confidence region and it contains the horizontal line $y = \pi(h)$ with a probability that is asymptotically smaller than $1-\alpha$. (In case of extreme variation of $(\delta_N(t))_{t\in[0,1]}$, it may even fail to contain *any* horizontal line.)

Constructing a confidence band on $(\delta_N(t))_{t\in[0,1]}$ (and therefore on the horizontal line $\pi(h)$) is nonetheless possible when using Donsker's theorem. This theorem states that the random functional

$$\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} \left(\frac{h(X_i) - \pi(h)}{\sqrt{V_\pi[h(X)]}} \right) \right\}_{t\in[0,1]}$$

converges in distribution (when N goes to infinity) to a standard Brownian motion on $[0,1]$, denoted $\{W(t)\}_{t\in[0,1]}$ (Billingsley, 1968). This is also the approximation used in the Kolmogorov-Smirnov test calibration.

Similarly, the random function

$$\left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} \{h(X_i) - \pi(h)\} \right)_{t\in[0,1]}$$

converges in distribution to a standard Brownian motion $\{W(t)\}_{t\in[0,1]}$. Donsker's theorem thus enables us to determine the right asymptotic setting to define a confidence band on the simulation output.

Therefore, given two bounding functions g and f on $[0,1]$ such that

$$\mathbb{P}\left(-g(t) \le W(t) \le f(t), t \in [0,1]\right) = 1-\alpha, \tag{4}$$

the random region

$$\left\{ (t,y) \in [0,1] \times \mathbb{R} \;\middle|\; y \in \left[\delta_N(t) - f(t)\frac{\sqrt{N\gamma_N(1)}}{\lfloor Nt \rfloor}, \delta_N(t) + g(t)\frac{\sqrt{N\gamma_N(1)}}{\lfloor Nt \rfloor} \right] \right\} \tag{5}$$

contains the horizontal line $\{\pi(h)\}_{t\in[0,1]}$ with an asymptotic probability equal to $1-\alpha$. Using Donsker's theorem hence leads to a simultaneous asymptotic confidence region.

We will derive the functions f and g based on a minimal area optimality criterion in Section 3, but we first consider below an alternative region that is more relevant for simulation purposes.

2.2 A predictive region

The confidence region (5) is not directly relevant for Monte Carlo approximation, because only the final estimate $\delta_N(1) = \widehat{\mathfrak{I}}_N$ should be considered for the estimation of $\pi(h)$. In particular, (5) is much less informative than the final (asymptotic) confidence interval based on the whole sample. More precisely, the interval

$$\left[\delta_N(1) - \frac{\sqrt{\gamma_N(1)}}{\sqrt{N}}\Phi^{-1}(1-\alpha/2), \delta_N(1) + \frac{\sqrt{\gamma_N(1)}}{\sqrt{N}}\Phi^{-1}(1-\alpha/2) \right] \tag{6}$$

contains $\pi(h)$ with a asymptotic probability of $1-\alpha$ and, due to the lack of simultaneity requirement, it is smaller than the slice of (5) for $t = 1$,

$$\left[\delta_N(1) - f(1)\frac{\sqrt{N\gamma_N(1)}}{N}, \delta_N(1) + g(1)\frac{\sqrt{N\gamma_N(1)}}{N} \right] .$$

Moreover, as pointed out previously, the quantity of interest from a Monte Carlo point of view is rather a prediction region that contains another Monte Carlo sequence with a given (asymptotic) probability. Given an iid sample $Y_1, \ldots, Y_n$ from π that is independent of $X_1, \ldots, X_n$, we can use the decomposition

$$\left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} (h(Y_i) - \delta_N(1)) \right)_{t\in[0,1]} = \left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} (h(Y_i) - \pi(h)) \right)_{t\in[0,1]} - \left(\frac{\lfloor Nt \rfloor\ \gamma_N(1)^{-1/2}}{\sqrt{N}} (\delta_N(1) - \pi(h)) \right)_{t\in[0,1]} ,$$

to deduce that it converges in distribution to a continuous random process that is the sum of a standard Brownian motion $\{W(t)\}_{t\in[0,1]}$ and of a random line $(tU)_{t\in[0,1]}$, U being a standard normal random variable independent of $\{W(t)\}_{t\in[0,1]}$.

Therefore, if we set $Z(t) = W(t) + tU$, $t \in [0,1]$, the prediction region for the Monte Carlo sequence can be reduced to the derivation of two bounds g and f on $[0,1]$ such that

$$\mathbb{P}(-g(t) \le Z(t) \le f(t), t \in [0,1]) = 1 - \alpha \,, \tag{7}$$

since the band

$$\left\{ (t,y) \in [0,1] \times \mathbb{R} \;\middle|\; y \in \left[\delta_N(1) - f(t)\frac{\sqrt{N\gamma_N(1)}}{\lfloor Nt \rfloor}, \delta_N(1) + g(t)\frac{\sqrt{N\gamma_N(1)}}{\lfloor Nt \rfloor} \right] \right\} \tag{8}$$

is a prediction region of level $1-\alpha$ for the sequence $\left(\sum_{i=1}^{\lfloor Nt \rfloor} h(Y_i) \Big/ \lfloor Nt \rfloor \right)_{t\in[0,1]}$.

As an aside, note that the process

$$\left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} \{h(Y_i) - h(X_i)\} \right)_{t\in[0,1]} = \left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} \{h(Y_i) - \pi(h)\} \right)_{t\in[0,1]} - \left(\frac{\gamma_N(1)^{-1/2}}{\sqrt{N}} \sum_{i=1}^{\lfloor Nt \rfloor} \{h(X_i) - \pi(h)\} \right)_{t\in[0,1]}$$

converges in distribution to a continuous random process which is the sum of two independent standard Brownian motions or, equivalently, the process $\{F(t)\}_{t\in[0,1]} = \left\{\sqrt{2}W(t)\right\}_{t\in[0,1]}$. Therefore, using the same bounds as in (4), the random region

$$\left\{(t,y)\in[0,1]\times\mathbb{R} \;\middle|\; y\in\left[\delta_N(t)-\sqrt{2}f(t)\frac{\sqrt{N\gamma_N(1)}}{\lfloor Nt\rfloor},\delta_N(t)+\sqrt{2}g(t)\frac{\sqrt{N\gamma_N(1)}}{\lfloor Nt\rfloor}\right]\right\}$$

is also a prediction region of level $1-\alpha$ for an independent Monte Carlo sequence. However, for each $t\in[0,1[$, the variability of $\delta_N(t)$ is more important than the one of $\delta_N(1)$ and, for the same Rao-Blackwell reason as above, the global variability of $\{\sqrt{2}W(t)\}_{t\in[0,1]}$ is also more important than the one of $\{W(t)+tU\}_{t\in[0,1]}$. There is thus little incentive in using the above prediction region, when compared with (8).

We will deal with the derivation of approximately optimal solutions to (7) in Section 4.

3 Minimal area confidence regions

As previously described, for computing simultaneous confidence regions of level $1-\alpha$, we have to derive two bounding functions g and f on $[0,1]$ such that (4) holds. Among the solutions to this problem, a natural optimality criterion is to select those such that the area of the confidence band $\int_0^1(f(x)+g(x))\,\mathrm{d}x$ is minimal. We are thus trying to solve a functional optimisation problem, namely to find the pair of functions u, v which together solve the problem

$$\min_{u,v}\int_0^1(u(t)+v(t))\mathrm{d}t$$

under the constraint that

$$\mathbb{P}(-v(t)\le W(t)\le u(t), t\in[0,1]) = 1-\alpha\,,$$

with $u\ge 0$ and $v\ge 0$.

From the symmetry of the standard Brownian motion, if the minimum is attained and is unique then we have $v=u$. So we can assume that the solution to the above problem satisfies $f=g$ and then replace the above problem with

$$\min_{u\ge 0}\int_0^1 u(t)\mathrm{d}t$$

under the constraint that $\mathbb{P}(-u(t)\le W(t)\le u(t), t\in[0,1]) = 1-\alpha$.

This minimization problem does not have a closed form solution and appears to be extremely difficult to solve. We will detail some local time approximations to its resolution, but let us first point out that the computation of the probability that a standard Brownian motion remains between two given boundaries is a complex question that has occupied many researchers for many years (see, for example, Anderson, 1960; Robbins and Siegmund, 1970; Durbin, 1971; Lerche, 1986; Durbin, 1992; Daniels, 1996; Borodin and Salminen, 2002; Li, 2003).

In this literature, explicit representations of two-sided boundary crossing probabilities are extremely rare and mostly address linear cases (Anderson, 1960; Hall, 1997), even though there exist a few nonlinear boundaries (Robbins and Siegmund, 1970; Daniels, 1996; Novikov et al., 1999). For most functions, though, computing two-sided boundary crossing probabilities for the Brownian motion requires some level of approximation.

3.1 Local time solution

Given the intractable difficulty of the derivation of the confidence band, we now switch to a related and much easier problem based on local times.

The original problem is to solve, for a fixed $0<\alpha<1$,

$$\min_u\left(\int_0^1 u(t)\mathrm{d}t\right) \qquad \text{subject to the constraint } \mathbb{P}\left\{-u(t)\le X(t)\le u(t) \text{ for all } t\in(0,1]\right\} = 1-\alpha\,,$$

where $X=\{X(t)\}_t$ is a continuous time process (in this section, X is a Wiener process $\{W(t)\}_t$ while, for the prediction problem, it is the modification $\{W(t)+tU\}_t$). Equivalently we can consider the dual problem

$$\min_u\left\{1-\mathbb{P}\left(-u(t)\le X(t)\le u(t) \text{ for all } t\in(0,1]\right)\right\} \qquad \text{subject to the constraint } \int_0^1 u(t)\mathrm{d}t=\beta\,,$$

where $\beta > 0$ is a given constant.

This formulation suggests the study of an alternative and easier problem, in which we replace the probability by a local time expectation:

$$\min_u \mathbb{E}\left[L^u(1) + L^{-u}(1)\right] \qquad \text{subject to the constraint } \int_0^1 u(t)\mathrm{d}t = \beta\,. \quad (9)$$

Here, $L^u(s)$ is the *local time* accumulated by X along the curve u up to time s, to be defined below, and we require u to be C^1-smooth.

This new problem is much easier because local time is additive. Therefore, firstly, we may restrict attention to the one-sided problem, that is, replace $\mathbb{E}[L^u(1) + L^{-u}(1)]$ by $\mathbb{E}[L^u(1)]$ and, secondly, it is then possible to localize the minimization problem in time, as now detailed.

Definitions of *local time* for Wiener processes and for other semimartingales can be found in Revuz and Yor (1999, Ch. VI); for our purposes we use the Itô-Tanaka formula (Revuz and Yor, 1999, Ch. VI, Theorem 1.2) which defines a local time L for a continuous semimartingale Y, started at 0, with increasing process $\langle Y, Y\rangle(t) = t$. (These conditions are equivalent to the requirement that Y be a Brownian motion with possibly non-stationary drift.) Applying the Itô-Tanaka formula to $X - u$, we obtain the following formula for the local time L^u of X at the time-varying level u (the C^1-smooth condition on u is required here):

$$\frac{1}{2}L^u(t) \quad = \quad \int_0^t u'(s)\mathbb{I}_{[X(s)>u(s)]}\mathrm{d}s + \int_0^t \mathbb{I}_{[X(s)\le u(s)]}\mathrm{d}X(s) - \min\{X(t), u(t)\}\,.$$

The local time $L^u(t)$ can be also interpreted as the time spent by X close to the curve u up to time t, in the sense that (see Revuz and Yor, 1999, Ch. VI, Corollary 1.9)

$$L^u(t) \quad = \quad \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^t \mathbb{I}_{[u(s)\le X(s)<u(s)+\varepsilon]}\mathrm{d}s\,. \quad (10)$$

Therefore, so long as X has a continuous transition probability density $p_{s,t}(x, y)$ (which is the density of $X(t)$ given $X(s) = x$), we can write

$$\mathbb{E}[L^u(1)] \quad = \quad \int_0^1 p_{0,s}(0, u(s))\mathrm{d}s\,.$$

For our purpose, we can assume furthermore that $X(s)$ is normally distributed with mean zero and variance $\tau(s)$. Hence the resolution of the minimization problem (9) amounts to solving

$$\min_u \left(\frac{1}{\sqrt{2\pi}}\int_0^1 \exp\left(-\frac{u(s)^2}{2\,\tau(s)}\right)\frac{\mathrm{d}s}{\sqrt{\tau(s)}}\right) \qquad \text{subject to the constraint } \int_0^1 u(t)\mathrm{d}t = \beta\,.$$

This minimization problem localizes: using a variation $u(s) + \varepsilon h(s)$ and differentiating with respect to ε, we find the condition for extremality to be

$$\frac{u(s)}{\sqrt{\tau(s)}}\exp\left(-\frac{1}{2}\left(\frac{u(s)}{\sqrt{\tau(s)}}\right)^2\right) \quad = \quad \kappa\tau(s)\,,$$

where κ is the variational constant (or Lagrange multiplier) connected to β. (In fact this integral minimization problem can be solved even if we suppose u to be merely measurable.) There are no solutions to this equation if $\kappa\tau(s) > 1/\sqrt{e}$, just one solution if equality holds, and two solutions (one above $\sqrt{\tau(s)}$, one below) if $\kappa\tau(s) < 1/\sqrt{e}$. A second variation argument shows that the minimization problem is solved by the larger solution if any exist, and otherwise we should take $u = 0$ (when $\kappa\tau(s) > 1/\sqrt{e}$).

Taking $\psi(a)$ to be the larger solution to $\psi e^{-\psi^2/2} = a$, if solutions exist, we see that the required minimizing u is given by

$$u^\star(s) \quad = \quad \begin{cases} \psi(\kappa\tau(s))\sqrt{\tau(s)} & \text{if } \kappa\tau(s) \le 1/\sqrt{e}\,, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

The function ψ can be expressed in terms of the so-called `ProductLog` or Lambert W function, the analytic solution $w(z)$ of $we^w = z$ (Corless et al., 1996). Indeed, selecting an appropriate branch of `ProductLog` (using the conventions of *Mathematica*), we find

$$\psi(a) \quad = \quad \sqrt{-\texttt{ProductLog}(-1, -a^2)} \text{ for } a \le 1/\sqrt{e}\,.$$

When X is the Brownian motion, $t(s) = s$ and the optimal solution $u^\star$ can be easily computed (either via mathematical software such as *Mathematica* or via an equivalent R function, available from the authors). For this type of functions, the resolution of the original problem then amounts to the computation of the appropriate constant κ to achieve a proper coverage level. For $\alpha = 0.05$, our simulations led to $\kappa = 0.105$ (see below for details).

3.2 Numerical results

Before detailing our resolution of the coverage approximation and the corresponding performances of some classes of solutions, we review some approximation techniques found in the literature. We denote by $P_1(u)$ the probability $\mathbb{P}\{|W(t)| \leq u(t), t \in [0,1]\}$.

Since the analytical resolution of the optimisation problem seems to be completely out of reach, we describe a numerical approach based on a partition $t_0 = 0 < t_1 < t_2 < \ldots < t_n = 1$ of the interval $[0,1]$ of size $n \geq 1$, with $\delta t_i = t_i - t_{i-1}$ and $\beta_i = u(t_i)$. We will distinguish below between three types of approximation, even though there exist many other possible approaches in the literature, like, e.g., polynomial and Poisson approximations, and Girsanov transformations.

The first type of approximation is based on the approach of Wang and Potzelberger (1997) and Potzelberger and Wang (2001). These authors show that, for any continuous function u,

$$P_1(u) = \lim_{n\to\infty} \mathbb{E}[\ell(W(t_1), W(t_2), \ldots, \ldots, W(t_n)]$$

where $\ell(x_1, \ldots, x_n)$ is defined by

$$\prod_{i=1}^{n} \mathbb{I}_{(-\beta_i < x_i < \beta_i)} \left[1 - \exp\left[-\frac{2}{\delta t_i}(-\beta_{i-1} - x_{i-1})(-\beta_i - x_i)\right] - \exp\left[-\frac{2}{\delta t_i}(\beta_{i-1} - x_{i-1})(\beta_i - x_i)\right]\right].$$

Using this result, Potzelberger and Wang (2001) derive the following Monte Carlo estimator of $P_1(u)$

$$\hat{P}_1(u)_{n,N} = \frac{1}{N}\sum_{j=1}^{N} \ell(W_j(t_1), W_j(t_2), \ldots, W_j(t_n))$$

where $W_1, \ldots, W_N$ are N independent standard Brownian motions on $[0,1]$. As n and N go to infinity, $\hat{P}_1(u)_{n,N}$ converges almost surely to $P_1(u)$. Under the assumptions that the boundary u is twice continuous differentiable with $u''(0) \neq 0$ and $u''(t) = 0$ at most in finitely many points $t \in]0,1]$, the authors proposed a special rule for choosing a sequence of "optimal partitions". Note that it is also possible to use a crude Monte Carlo method to approximate $P_1(u)$, that is to estimate $P_1(u)$ with

$$\widetilde{P_1(u)}_{n,N} = \frac{1}{N}\sum_{j=1}^{N}\prod_{i=0}^{n} \mathbb{I}_{\{-\beta_i \leq W_j(t_i) \leq \beta_i\}}.$$

However, the variance of the Monte Carlo estimator of Potzelberger and Wang (2001) is inevitably smaller than the one of a crude Monte Carlo method insofar as $\ell(x_1, \ldots, x_n) \leq 1$ for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

The second type of approximation is due to Novikov et al. (1999), based on the representation

$$P_1(u) = \mathbb{E}\left[\prod_{i=0}^{n-1} p_i(u|W(t_i), W(t_{i+1}))\right],$$

where $p_i(u|x,y) = \mathbb{P}(-u(t) < W(t) < u(t), t_i \leq t \leq t_{i+1}|W(t_i) = x, W(t_{i+1}) = y)$. Let u_n be the piecewise linear approximation of u with connecting points $(t_i, u(t_i)), \quad i = 0, \ldots, n$. Hall (1997) calculated the conditional distribution of crossing the upper and lower linear boundaries. Using this result, Novikov et al. (1999) have built a fast algorithm that calculates $P_1(u_n)$. Finally, as $u_n(t)$ goes to $u(t)$ uniformly on $[0,1]$, it follows from the continuity property of probability measures that $\lim_{n\to\infty} P_1(u_n) = P_1(u)$. As n tends to infinity, the algorithm of Novikov et al. (1999) gives a convergent approximation of $P_1(u)$.

For the third type of approximation, Durbin (1971), Park and Schuurmann (1976), Loader and Deely (1987) have proposed to use some Volterra integral equations. Various integral equations involving the distribution of

Class	$u(t)$	minimum	$\int_0^1 u(t)\mathrm{d}t$
1	a	$a = 2.242$	2.242
2	$a + bt$	$a = 1,\ b = 1.77$	1.885
3	$a + b\sqrt{t}$	$a = 0.3,\ b = 2.35$	1.866
$\star$	$u^\star(t)$	$\kappa = 0.105$	1.865

Table 1: Optimal solutions for three classes of parametrised boundaries, along with the local time solution u^* (the two last lines are based on $5,000$ simulated Brownian paths with partition size 500).

the first exit time from a corridor between two given boundaries have been published. Based on the discretization of integral equations, Loader and Deely (1987) have given a numerical solution that is very efficient and we will use this approximation.

In our numerical experiment, besides the solution given by the local time approximation, we have only considered three classes of functions: the constant class $u(t) = a$ $(a > 0)$, the linear class $u(t) = a+bt$ $(a, b > 0)$, the square root class $u(t) = a+b\sqrt{t}$ $(a, b > 0)$. For each class, we have derived parameters such that $P_1(u) = 0.95$ and $\int_0^1 u(t)\mathrm{d}t$ is minimum. While the two first classes allow for analytical expressions for boundary crossing probabilities, in the case of the third class, we have used the approximation method of Loader and Deely (1987). Table 1 presents the results of this numerical experiment. Figure 1 presents 5000 simulated Brownian paths together with the solutions in each class.

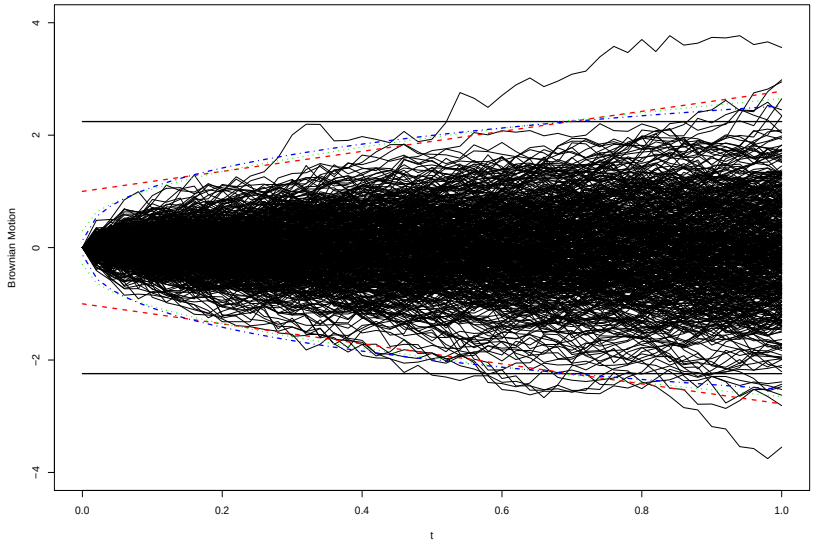

Figure 1: 5000 Brownian paths and functions u such that $P_1(u) = 0.95$, as given by Table 1. The two last functions in Table 1 are almost identical.

As clear from both Figure 1 and Table 1, the gain reached in increasing the complexity of u is very marginal: moreover while the local time approximation function $u^\star$ actually does better that the approximations studied here, nevertheless the difference is very small. We also ran simulations for other classes of functions, such as $u(t) = a + bt + ct^2 + d\log(t)$, with no visible improvement. We thus propose to use either $u^\star$ or the function $u(t) = 0.3 + 2.35\sqrt{t}$ (which is obviously simpler to compute) for constructing a confidence region of level 0.95.

4 Minimal area prediction regions

The construction of prediction regions of level $1-\alpha$ is very similar to the above, with the difference that the limiting process is now $\{W(t)+tU\}_t$. The optimisation problem is thus

$$\min_{u\geq 0} \int_0^1 u(t)\mathrm{d}t$$

under the constraint that $P_2(u) = \mathbb{P}(-u(t) \leq W(t)+tU \leq u(t), t\in[0,1]) = 1-\alpha$.

Similar difficulties beset the derivation of the optimal bound and we use once more approximate boundaries to evaluate the boundary-crossing probability, either derived from the local time representation or by "crude" Monte Carlo methods.

4.1 Local time solution

The local time approximation of Section 3.1 also applies to this setting, in the sense that, for the modified problem (9), we have an explicit solution given in (11). The difference with the above section is that, as we are dealing with $X(t) = W(t)+tU$, we have $\tau(s) = s(1+s)$, and so the extremal envelopes are now given by

$$u^\star(s) \quad = \quad \begin{cases} \psi(\kappa s(1+s))\sqrt{s(1+s)} & \text{for } \kappa s(1+s) \leq 1/\sqrt{e}\,, \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

for varying κ.

Once again, these optimal solutions $u^\star$ can be computed numerically (for example in *Mathematica*). By way of example, Figure 2 shows upper and lower envelopes $\pm u^\star$ at $\kappa = 1/20$ together with 100 simulations of the Brownian motion with empirical mean, $W(t)+tU$.

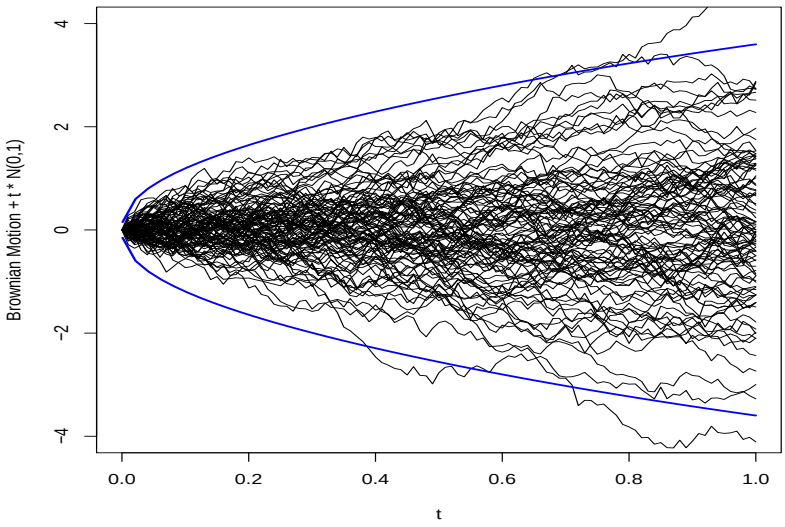

Figure 2: 100 Brownian/empirical mean trajectories together with a two-sided version of the extremal $u^\star$ for $\kappa = 1/20$.

For our determination of the $1-\alpha$ confidence region, we must thus calibrate κ to achieve a given coverage. Using the simulation experiment detailed below, we found $\kappa = 0.095$ for $\alpha = .05$.

4.2 Numerical results

Since we do not have a Volterra expression as for the Brownian motion, we use a cruder Monte Carlo approach to derive coverage probabilities and optimal values of parameterised boundaries.

Once more, we use a partition $t_0 = 0 < t_1 < t_2 < \ldots < t_n = 1$ of the interval $[0,1]$ of size $n \geq 1$, with $\beta_i = u(t_i)$. It follows from the continuity property of probability measures that, if $u_n(t) \longrightarrow u(t)$ uniformly on $[0,1]$, then

$$\lim_{n\to\infty} P_2(u_n) = P_2(u) \quad \text{and} \quad \lim_{n\to\infty} \prod_{i=1}^{n} \mathbb{P}(-\beta_i \leq W(t_i) \leq \beta_i) = P_2(u)\,.$$

Moreover, the product of the $\mathbb{P}(-\beta_i \leq W(t_i) \leq \beta_i)$ can be estimated by a Monte Carlo approximation, namely

$$\hat{P}_2(u)_{n,N} = \frac{1}{N} \sum_{j=1}^{N} \prod_{i=0}^{n} \mathbb{I}_{-\beta_i \leq W_j(t_i) \leq \beta_i}$$

where $W_1, \ldots, W_N$ are N independent standard Brownian motions on $[0,1]$. As N goes to infinity, $\hat{P}_2(u)_{n,N}$ converges almost surely to $\prod_{i=1}^{n} \mathbb{P}(-\beta_i \leq W(t_i) \leq \beta_i)$ and, as n and N both go to infinity, $\hat{P}_2(u)_{n,N}$ is a convergent estimator of $P_2(u)$.

For the same three parameterised classes of functions $u(\cdot)$ as above, we have evaluated the function u that minimises its subgraph under the coverage constraint $P_2(u) = 0.95$. Table 2 presents the results, along with the corresponding local time solution. Figure 3 plots $5,000$ paths $W(t) + tU$ against the four functions.

Class	$h(t)$	Minimum	$\int_0^1 h(t)\mathrm{d}t$
1	a	$a = 2.94$	2.94
2	$a + bt$	$a = 1.2$, $b = 2.1$	2.25
3	$a + b\sqrt{t}$	$a = 0.1$, $b = 3.15$	2.2
$\star$	$u^\star(t)$	$\kappa = .095$	2.189

Table 2: Optimal solutions for three classes of parametrised boundaries, along with the local time solution u^*, based on $1,000,000$ simulated Brownian paths with partition size 1000.

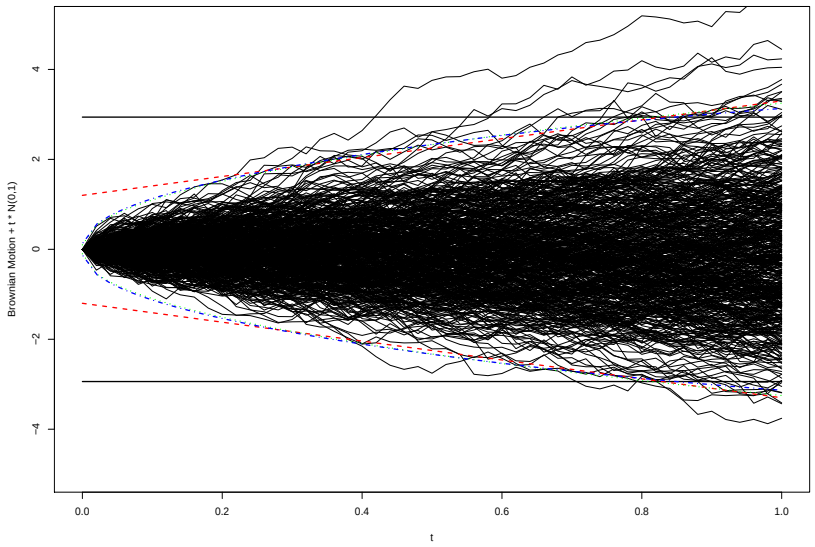

Figure 3: 5000 $W(t) + tU$ paths and functions h such that $P(h) = 0.95$

Once again, the gain brought by the local time solution or other classes of boundary functions, like $u(t) = a + bt + ct^2 + d\log(t)$, is quite minimal when compared with the basic function $u_0(t) = 0.1 + 3.15\sqrt{t}$. In practice, since u_0 is almost indistinguishable from $u^\star$, it seems more convenient to use u_0 for the evaluation of Monte Carlo variability.

5 Example

We consider a toy problem when π is $\mathcal{N}(0,1)$ and $h(x) = \exp(x^2/4.01)$. The reason why we picked this specific function h is that it provides a borderline for an infinite variance estimator—had we chosen 4.0 instead of 4.01 in the denominator, the regular Monte Carlo estimator would have had an infinite variance and the CLT would not apply). The Monte Carlo estimator of $\pi(h)$ is

$$\delta_N(1) = \frac{1}{n}\sum_{i=1}^{N} \exp\left(X_i^2/4.01\right)$$

where $X_1, \ldots, X_n$ is an iid $\mathcal{N}(0,1)$ sample. The average behaviour of the sequence of $\delta_N(t)$'s is illustrated by Figure 4, for 1000 replicas of $N = 1000$ simulations. Note the interesting feature that one realisation out of the 1000 replicas takes a huge jump around $t = 0.55$: this kind of behaviour is typical of estimators with infinite (or almost infinite) variances. A correct predictive confidence band should not include this extreme behaviour; however, were we to run a single simulation and observe this type of jump then the resulting confidence band should be to some extent enlarged. The repetition of the Monte Carlo sequences also shows a $1/\sqrt{t}$ shape that validates the application of the CLT.

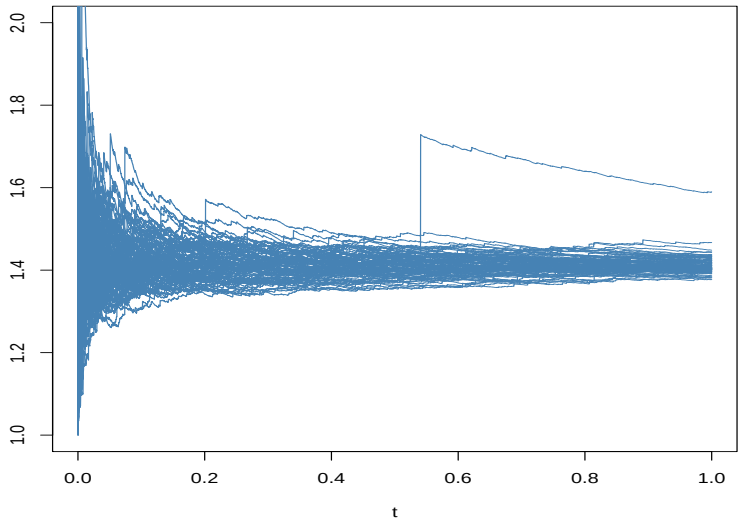

Figure 4: 1000 replicas of a sequence of Monte Carlo estimators $\delta_N(t)$ for $N = 1000$.

As noted in Section 2, the use of a point-wise normal confidence interval on the path-wise simulation is in principle incorrect in terms of coverage. In addition, as shown by Figure 5, it suffers from the practical defect that the confidence band has boundaries that reflect the estimator path (with a $1/\sqrt{t}$ range decrease). Moreover, although this particular issue does not arise here, in the event of several big jumps of the estimator along iterations, it may even occur that the confidence band develops discontinuities. Furthermore, when comparing the scale of the band with Figure 4, which demonstrates a typical range of variation, we can see that it is much too small to accommodate the variation of the majority of the sequences.

If we now consider Figure 6, based on the same realisation of $\{\delta_{1000}(t)\}_{t\in[0,1]}$ as in Figure 5, the simultaneous confidence region based on (5) is much more satisfactory in that, while its variations still track the non-monotonic variation of the estimator sequence, and thus suffers from a lack of regularity that is inappropriate for a confidence band, the range is wider, as shown in Figure 7, and thus tends to erase the jumps in the estimator itself.

If we now turn to the most relevant confidence band, as derived in Section 4, namely (8), it is represented in Figure 8 for the same realization of the Monte Carlo sequence as in Figure 5. The difficulties of the previous bands have disappeared in that the confidence band is centred at the final value of the estimate and do not present irregular jumps on their boundary. Intuitively, this band should be wider than the previous ones, in

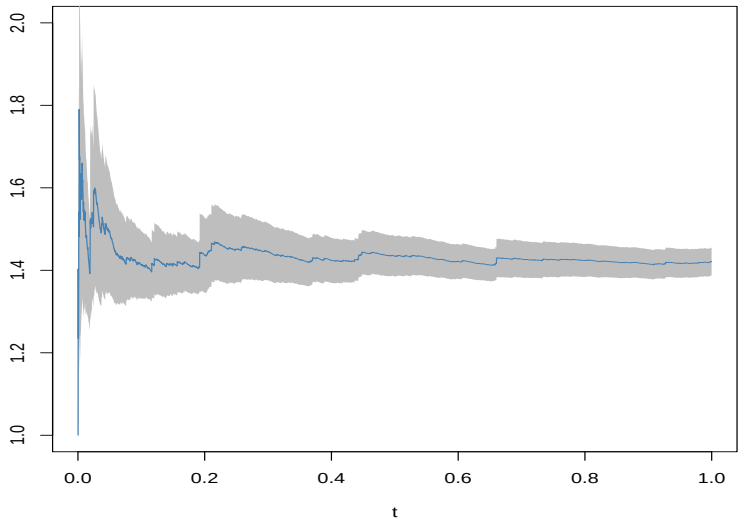

Figure 5: Confidence band at level 0.95 derived from a naïve application of the CLT for one sequence of Monte Carlo estimators represented by the inner curve.

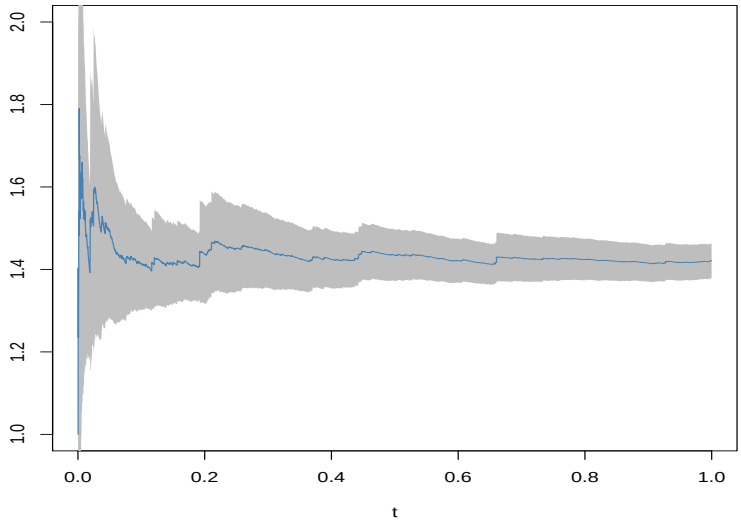

Figure 6: Confidence band at level 0.95 derived from (5) for the same realisation as in Figure 5.

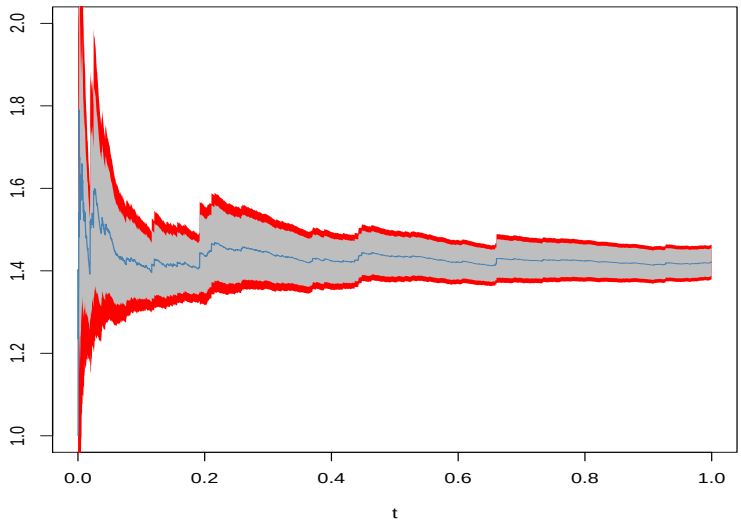

Figure 7: Comparison of the confidence bands of Figures 5 (tighter) and 6 (wider).

order to accommodate variations in the whole sequence but this is not exactly the case, as shown in Figure 9 by the comparison between the three bands. While (8) leads on average to a wider region, there are occurrences when (5) exceeds the boundaries of (8); this relates to the defect of (5) concerning discontinuous jumps, as mentioned above.

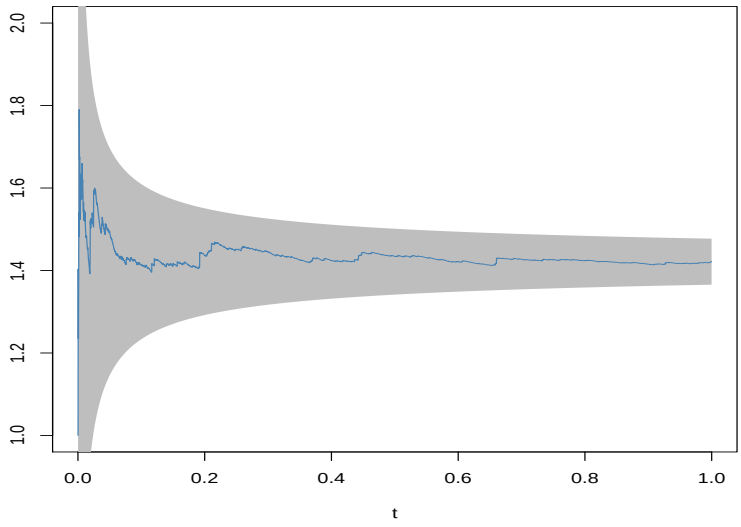

Figure 8: Confidence band at level 0.95 derived from (5) for the same realisation as Figure 5.

6 Conclusion

It seems to us that this assessment should be used for every Monte Carlo experiment, given that the bound u has only to be determined once for a confidence level α. Indeed, once we have determined a proper boundary

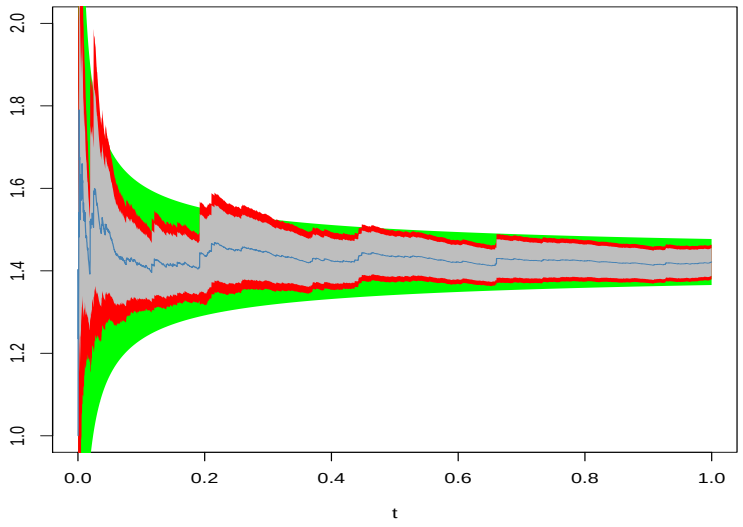

Figure 9: Comparison of the confidence bands of Figures 5, 6 and 8.

function $u^\star$, available as an R program from the authors, there is no cause for computational worry since the additional cost is completely negligible. We also stress that the extension to Markov Chain Monte Carlo settings causes no especial difficulty so long as the *same* functional CLT applies.

Acknowledgements

The first author is grateful to Sigurd Assing and to Larbi Alili for helpful conversations about boundary problems. The second author is grateful to Catherine Loader for sending him R programs related to the computation of the boundary crossing probabilities, as well as for useful discussions. The third author is grateful to both the Institute for Mathematical Sciences, National University of Singapore, for inviting him to attend the workshop on Markov Chain Monte Carlo in March 2004, thus giving him the opportunity to collaborate with the first author, and to George Casella for discussions about a preliminary perspective on that problem that eventually led to Chapter 4 in Robert and Casella (2004). Feedback from the audience of the ISBA 2004 poster session in Viña del Mar, Chile, is also gratefully acknowledged. A conversation with Marc Yor on theoretical foundations was equally helpful.

References

Anderson, T. (1960). A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size. *Ann. Math. Statist.*, 31(1):165–197.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

Billingsley, P. (1995). *Probability and Measure*. John Wiley, New York, third edition.

Borodin, A. and Salminen, P. (2002). *Handbook of Brownian motion - facts and formulae*. Birkhauser Verlag, 2 edition.

Corless, R., Gonnet, G., Hare, E., Jeffrey, D., and Knuth, D. (1996). On the Lambert W function. *Adv. Computational Math.*, 5:329–359.

Daniels, H. (1996). Approximating the first crossing-time density for a curved boundary. *Bernoulli*, 2:133–143.

Durbin, J. (1971). Boundary crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the kolmogorov-smirnov test. *J. Appl. Prob.*, 8:431–453.

Durbin, J. (1992). The first-passage density of the Brownian motion process to a curved boundary. *J. Appl. Prob.*, 29:291–304.

Hall, W. (1997). The distribution of Brownian motion on linear stopping boundaries. *Sequential Anal.*, 4:345–352.

Lerche, H. R. (1986). *Boundary Crossing of Brownian Motion.* Lecture Notes in Statistics. Springer-Verlag.

Li, W. V. (2003). The first exit time of a Brownian motion from an unbounded convex domain. *Ann. Prob.*, 31(2):1078–1096.

Loader, C. and Deely, J. (1987). Computations of Boundary Crossing Probabilities for the Wiener Process. *J. Stat. Comp. Simul.*, 27:96–105.

Novikov, A., Frishling, V., and Kordzakhia, N. (1999). Approximation of boundary crossing probabilities for a Brownian motion. *J. Appl. Prob.*, 36:1019–1030.

Park, C. and Schuurmann, F. (1976). Evaluation of barrier crossing probabilities for Wiener path. *J. Appl. Prob.*, 17:197–201.

Potzelberger, K. and Wang, L. (2001). Boundary crossing probability for Brownian motion. *J. Appl. Prob.*, 38:152–164.

Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition.

Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the Wiener process and partials sums. *Ann. Math. Statist.*, 38:152–164.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods.* Springer-Verlag, second edition.

Wang, L. and Potzelberger, K. (1997). Boundary crossing probability for Brownian motion and general boundaries. *J. Appl. Prob.*, 34:54–65.

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399